GxS

CLC **Genomics** Server
End User

USER MANUAL

Manual for
CLC Genomics Server 6.0
Windows, Mac OS X and Linux

February 7, 2014

This software is for research purposes only.

CLC bio, a QIAGEN Company
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark

Contents

Chapter 1

Introduction

The *CLC Server* is the central part of CLC bio's enterprise solutions. You can see an overview of the server solution in figure 1.1).

The basic idea is that you can have the Server store data and run jobs centrally and thereby offload personal and typically smaller computers. For the user, the difference between working just with a Workbench and working with a Workbench *and* a server is very small. All the mechanisms for managing data, using the tools and visualizing the data are the same.

This user manual is intended to show how to take advantage of the server as a user. For information about administrating the server, please see the Administrator Manual.

The *CLC Genomics Server* is shipped with the following tools and analyses that can be started from the *CLC Genomics Workbench* or *CLC Server Command Line Tools*:

- Import

- Export

- Download Genome

- Alignments and Trees

    - Create Alignment
    - Create Tree
    - Maximum Likelihood Phylogeny

- General Sequence Analysis

    - Extract Annotations
    - Extract Sequences

- Nucleotide Analysis

    - Translate to Protein
    - Convert DNA to RNA
    - Convert RNA to DNA
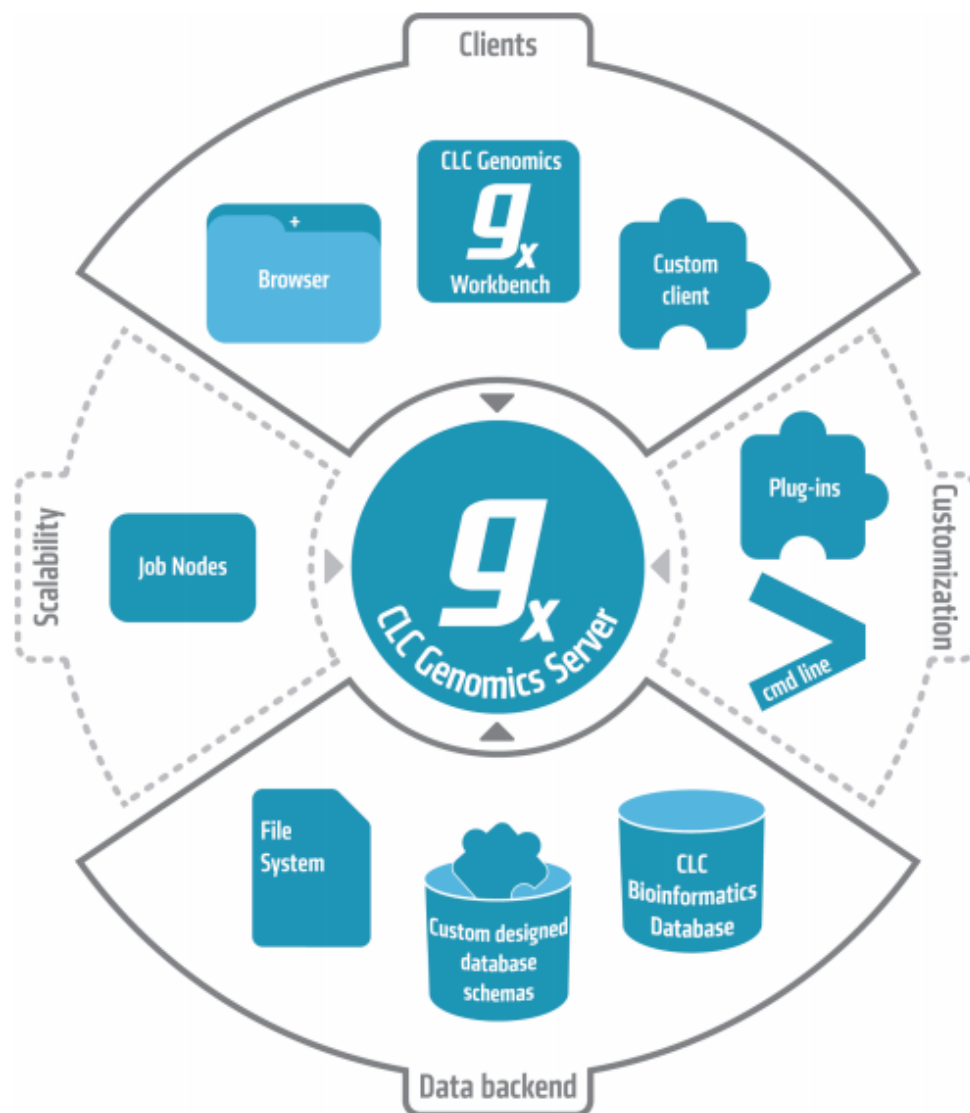   - Reverse Complement Sequence

Figure 1.1: *An overview of the server solution from CLC bio. Note that not all features are included with all license models.*

  - – Reverse Sequence
  - – Find Open Reading Frames
- Sequencing Data Analysis
  - – Trim Sequences
  - – Assemble Sequences
  - – Assemble Sequences to Reference
  - – Secondary Peak Calling
- Primers and Probes
  - – Find Binding Sites and Create Fragments
- BLAST

- – BLAST
- – BLAST at NCBI
- – Download BLAST Databases
- – Create BLAST Database
- – Manage BLAST Databases

- NGS Core Tools

  - – Create Sequencing QC Report
  - – Merge Overlapping Pairs
  - – Trim Sequences
  - – Map Reads to Reference
  - – Local Realignment
  - – Create Detailed Mapping Report
  - – Merge Read Mappings
  - – Extract Consensus Sequence
  - – Process Tagged Sequences (Multiplexing)

- Track Tools

  - – Convert to Tracks
  - – Convert from Tracks
  - – Merge Annotation Tracks
  - – Annotate with Overlap Information (Annotate and Filter)
  - – Extract Reads Based on Overlap (Annotate and Filter)
  - – Filter Annotations on Name (Annotate and Filter)
  - – Filter Based on Overlap (Annotate and Filter)
  - – Create GC Content Graph Tracks (Graphs)
  - – Create Mapping Graph Tracks (Graphs)
  - – Identify Graph Threshold Areas(Graphs)

- Resequencing Analysis

  - – Create Statistics for Target Regions
  - – Probabilistic Variant Detection
  - – Quality-based Variant Detection
  - – InDels and Structural Variants
  - – Coverage Analysis
  - – Annotate from Known Variants (Annotate and Filter Variants)
  - – Filter against Known Variants (Annotate and Filter Variants)
  - – Annotate with Exon Numbers (Annotate and Filter Variants)
  - – Annotate with Flanking Sequences (Annotate and Filter Variants)
  - – Filter Marginal Variant Calls (Annotate and Filter Variants)

- **–** Filter Reference Variants (Annotate and Filter Variants)
- **–** Compare Sample Variant Tracks (Compare Variants)
- **–** Compare Variants within Group (Compare Variants)
- **–** Fisher Exact Test (Compare Variants)
- **–** Trio Analysis (Compare Variants)
- **–** Filter against Control Reads (Compare Variants)
- **–** GO Enrichment Analysis (Functional Consequences)
- **–** Amino Acid Changes (Functional Consequences)
- **–** Annotate with Conservation Score (Functional Consequences)
- **–** Predict Splice Site Effect (Functional Consequences)

- Transcriptomics Analysis/ Expression Analysis

  - **–** RNA-Seq Analysis
  - **–** Extract and Count (Small RNA Analysis)
  - **–** Annotate and Merge Counts (Small RNA Analysis)
  - **–** Create Histogram (General Plots)

- Epigenomics Analysis

  - **–** ChIP-Seq Analysis

- De Novo Sequencing

  - **–** De Novo Assembly
  - **–** Map Reads to Contigs

The functionality of the *CLC Genomics Server* can be extended by installation of Server plugins. The available plugins can be found at http://www.clcbio.com/server_plugins.

Chapter 2

Using the server from a CLC Workbench

2.1 Installing relevant plugins in the Workbench

In order to use the *CLC Genomics Server* from a CLC Workbench, you need to install the CLC Workbench Client Plugin in the Workbench. This will allow you to log into the CLC Server, access data from the CLC Server data locations and submit analyses to your CLC Server.

Plugins are installed using the Plugins and Resources Manager[1], which can be accessed via the menu in the Workbench

> **Help** | **Plugins and Resources (🖼)**

or via the **Plugins (🖼)** button on the Toolbar.

From within the Plugins and Resources Manager, choose the Download Plugins tab and click on the CLC Workbench Client Plugin. Then click in the button labeled **Download and Install**.

If you are working on a system not connected to the network, then you can also install the plugin by downloading the cpa file from the plugins page of our website

http://www.clcbio.com/clc-plugin/

Then start up the Plugin manager within the Workbench, and click on the button at the bottom of the Plugin manager labeled **Install from File**.

You need to restart the Workbench before the plugin is ready for use.

2.2 Log into the server

Once the plug-in is installed, log into the server:

> **File** | **CLC Server Login (S)**

That will bring up a login dialog as shown in figure 2.1).

The first time you log into the server, you have to expand the dialog by clicking **Advanced**. That will allow you to enter the host and port for the server as shown in figure 2.2).

---

[1]In order to install plug-ins on many systems, the Workbench must be run in administrator mode. On Windows Vista and Windows 7, you can do this by right-clicking the program shortcut and choosing "Run as Administrator".
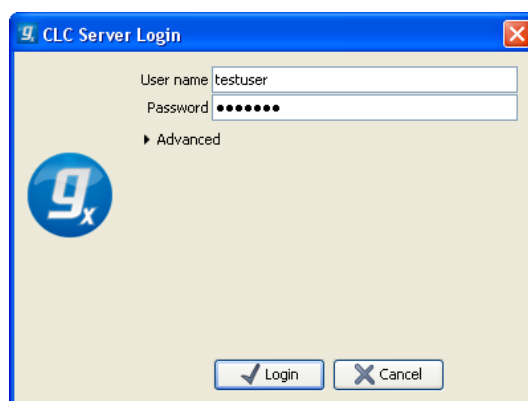
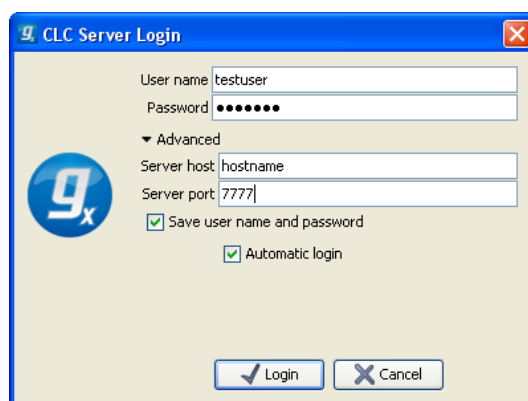Figure 2.1: *Expand the login dialog by clicking Advanced.*

Figure 2.2: *Specifying host and port.*

In addition you can choose to save user name and password and automatically log into the server when the Workbench starts.

Note that you need to get the login information from your server administrator.

When you press **Login**, the Workbench connects to the server. You will see a progress bar in the login dialog. If the login is successful, the dialog will disappear, and you will be able to use the server as described below.

2.3   Browsing and searching data from the workbench

Once you are logged in, the data locations on the server are shown in the **Navigation Area** (Figure 2.3).

Once logged in, the server data locations can be used in the same way as local data locations. We refer to the users manual of the Workbench for information about using the **Navigation Area** (click the location and press **F1** on the keyboard to get context help).

You may also wish to have a look at the search chapter, *Searching your data*, in the workbench user manual (press F1 and look for *Searching your data*).

It is possible to have both local and server locations added at the same time. This means that you can work on e.g. temporary sequences located on your own computer and then when you have more complete results, you can drag the elements to a folder in the server location.
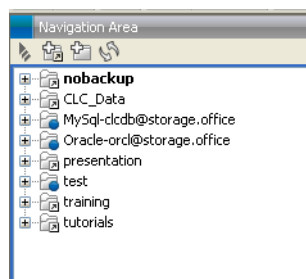
Figure 2.3: *Three server locations on the server appears in the Navigation Area (marked with blue dots). The remaining five local folders are file locations in the Workbench that are independent of the server.*

2.3.1 Deleting data

When you delete data located on the server, it will be placed in the **Recycle bin** (🗑) in the same way as when you delete data located on the Workbench. The data in the recycle bin can only be accessed by you and the server administrator.

Please note that the server administrator might have configured the recycle bin to be automatically emptied at regular intervals.

2.4   Running analyses on the server

The tools available on the server are listed in the introduction (see section 1).

For more information on the tools, please see the manual for *CLC Genomics Workbench* at http://www.clcbio.com/usermanuals or click the **Help** (?) button in the dialogs that are shown when you run the tools.

When you run the analyses you will be faced with an initial dialog with two choices:

- **Workbench**. This will run the analysis on your computer.

- **Server**. This will run the analysis on the server.

An example of such a dialog is shown in figure 2.4.

Note that you can check the **Remember setting and skip this step** option to always use the selected option when the analysis dialog is shown. If you change your mind later on and want to switch, simply click **Previous** in the dialog and you will be able to change the setting.

Running the analyses on the server is identical to running the same analyses in the Workbench with three exceptions:

- You can only select data from the data locations connected to the server. This means that the **Navigation Area** is filtered to only show these locations.

- You have to save the result. In the last step, you can normally choose between **Open** and **Save**, but for the analyses performed on the server, you have to choose **Save**.

- When you click **Finish**, the analysis is performed on the server. This means that you can close the Workbench or disconnect the computer from the server, and the analysis will
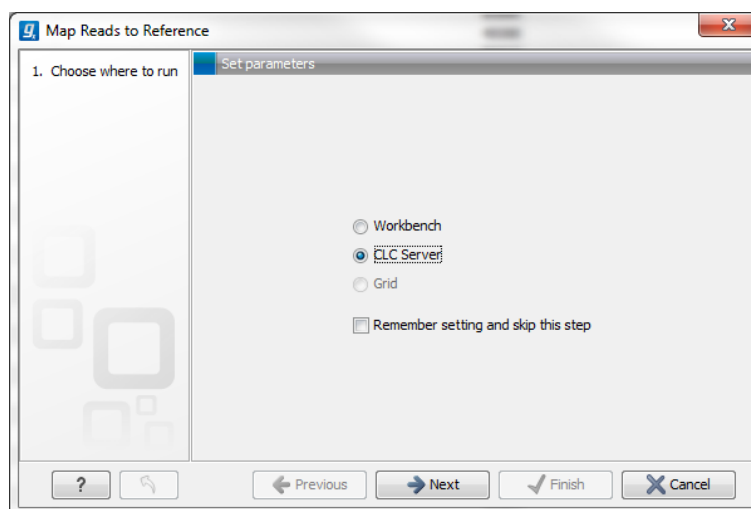
Figure 2.4: *Selecting where to run the analysis.*

continue to run. When you log in after the analysis is done, you will get a notification in the Workbench.

Note that when using the import tool (Figure 2.5) for importing high-throughput sequencing data on the server (only for Genomics Workbench and Server), the first part of the import is uploading the data to the server and during this part, the Workbench has to maintain connection to the server. If you try to close the Workbench during upload, you will get a warning dialog. The upload runs as a local process in **Processes** tab in the lower left corner, and when the upload is done, a new process is started that has a server icon and runs on the server.
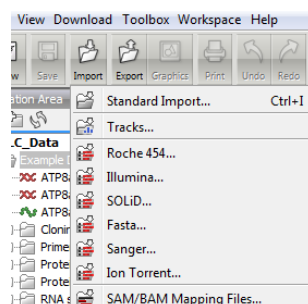

Figure 2.5: *The import tool in the Workbench.*

If you need help on the analyses themselves, please click the **Help** (?) button in the dialog. That will take you to the relevant part of the user manual of the Workbench.

2.5   Accessing files on, and writing to, areas of the server filesystem

There are circumstances when it is beneficial to be able to interact with (non-CLC) files directly on your server filesystem. A common circumstance would be importing high-throughput sequencing data from folders where it is stored on the same system that your *CLC Genomics Server* is running on. This could eliminate the need for each user to copy large sequence data files to the machine their CLC Workbench is running on before importing the data into a *CLC Genomics Server* CLC server data area. Another example is if you wish to export data from CLC format to other formats and save those files on your server machine's filesystem (as opposed to saving the files in the system your Workbench is running on).

From the administrator's point of view, this is about configuring folders that are safe for the *CLC Genomics Server* to read and write to on the server machine' system.

This means that users logged into the *CLC Genomics Server* from their Workbench will be able to access files in that area, and potentially write files to that area. Note that the *CLC Genomics Server* will be accessing the file system as the *user running the server process* - not as the user logged into the Workbench. This means that you should be careful when opening access to the server filesystem in this way. Thus, only folders that do not contain sensitive information should be added.

Folders to be added for this type of access are configured in the web administration interface **Admin** tab. Under **Main configuration**, open the **Import/export directories** (Figure 2.6) to list and/or add directories.
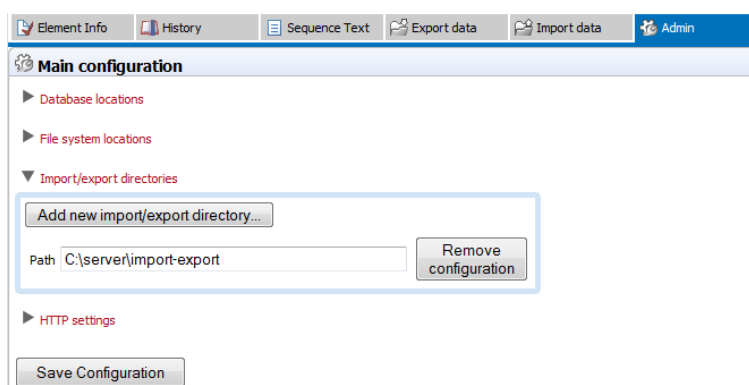

Figure 2.6: *Defining source folders that should be available for browsing from the Workbench.*

Press the **Add new import/export directory** button to specify a path to a folder on the server. This folder and all its subfolders will then be available for browsing in the Workbench for certain activities (e.g. importing data functions).

The import/export directories can be accessed from the Workbench via the Import function in the Workbench. If a user, that is logged into the *CLC Genomics Server* via their CLC Workbench, wishes to import e.g. high throughput sequencing data, an the option shown in figure 2.7 will appear.
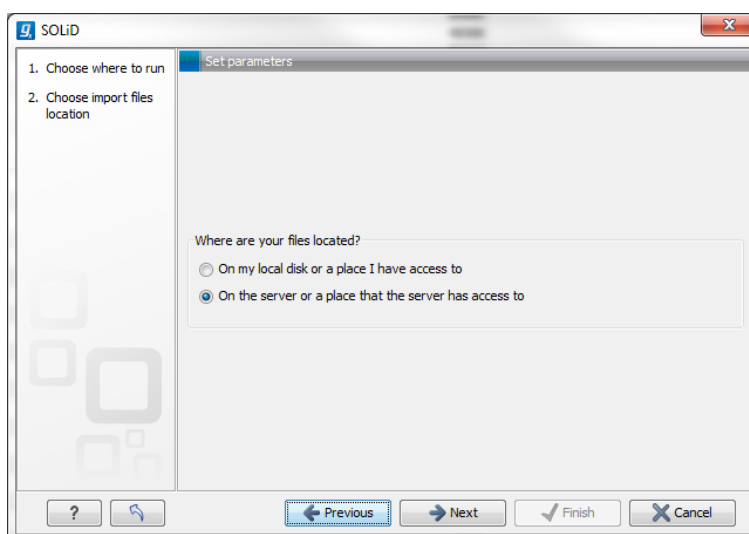

Figure 2.7: *Deciding source for high-throughput sequencing data files.*

On my local disk or a place I have access to means that the user will be able to select files from the file system of the machine their CLC Workbench is installed on. These files will then be transferred over the network to the server and placed as temporary files for importing. If the user chooses instead the option *On the server or a place the server has access to*, the user is presented with a file browser for the selected parts of the server file system that the administator has configured as an Import/export location (Figure 2.8).
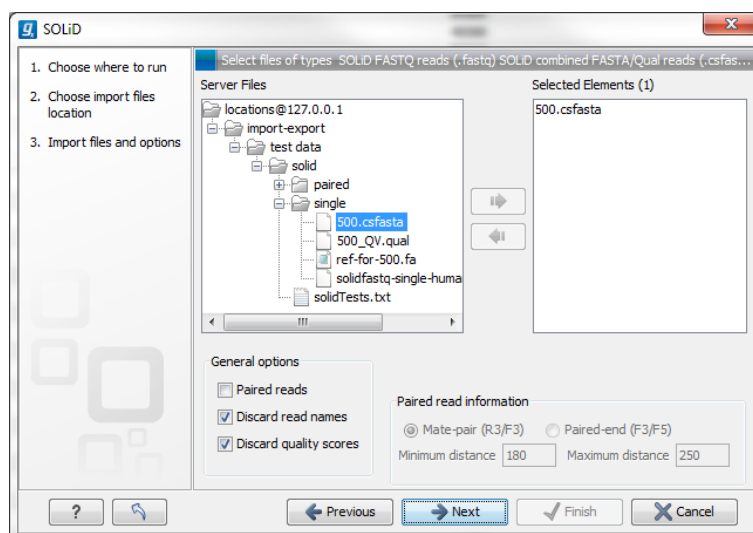

Figure 2.8: *Selecting files on server file system.*

Note: Import/Export locations should NOT be set to subfolders of any defined CLC file or data location. CLC file and data locations should be used for CLC data, and data should only be added or removed from these areas by CLC tools. By definition, an Import/Export folder is meant for holding non-CLC data, for example, sequencing data that will be imported, data that you export from the Genomics Server, or blast databases. Note that your server administrator needs to configure the server to import files directly from the server file system.

2.6 Monitoring processes

Just as for processes running locally, you will be able to monitor processes running on the server. Click the **Processes** tab next to the **Toolbox** and you will see the list of processes (see figure 2.9).

Note that the processes running on the server have a server icon (**S**) or (**S**) whereas the processes running locally have specific icons like ( ) and ( ). You can also see that two of the server processes are queued.

Contrary to processes running locally, running server processes will appear even if you restart the Workbench (and log into the server). Server processes that are finished when you close the Workbench will not be shown when you start the Workbench again.
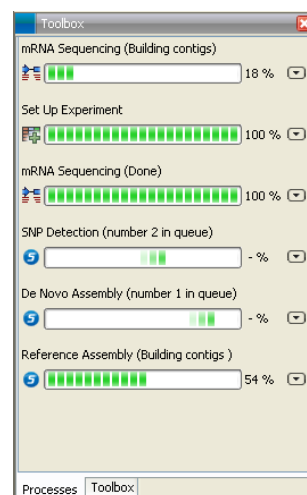
Figure 2.9: *Monitoring processes.*

Chapter 3

Using a web browser as client

Besides using the CLC Workbench as client, you also have access to the server from the web interface. The web interface gives you access to browsing and searching data and importing and exporting data.

Simply type the address of the server into your browser, followed by the port number, and you will see the login dialog (see figure 3.1) (you will need to get information from your server administrator about the server address).
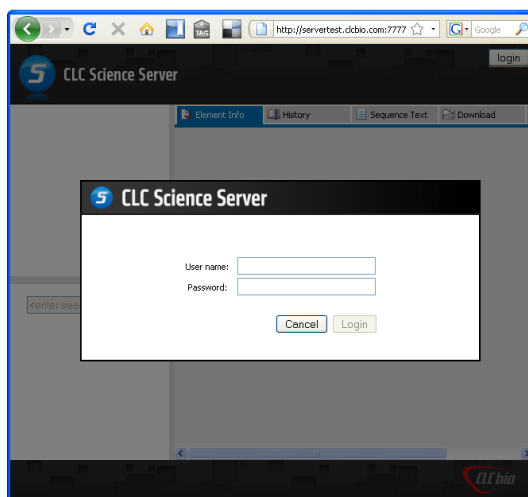

Figure 3.1: *The web interface of the server.*

3.1 Browsing and searching data

To the left, you will see the data locations connected to the server. You can browse the folder hierarchy of the data location. When you click an element in the tree, a number of options are available:

- Click the **Element Info** ( ) tab to see the properties of this element. Note that you can edit the information in this view.

- Click the **History** ( ) tab to see the history of this element.

- Click the **Sequence Text** (▤) tab to see a textual representation of this element (only works for sequences)

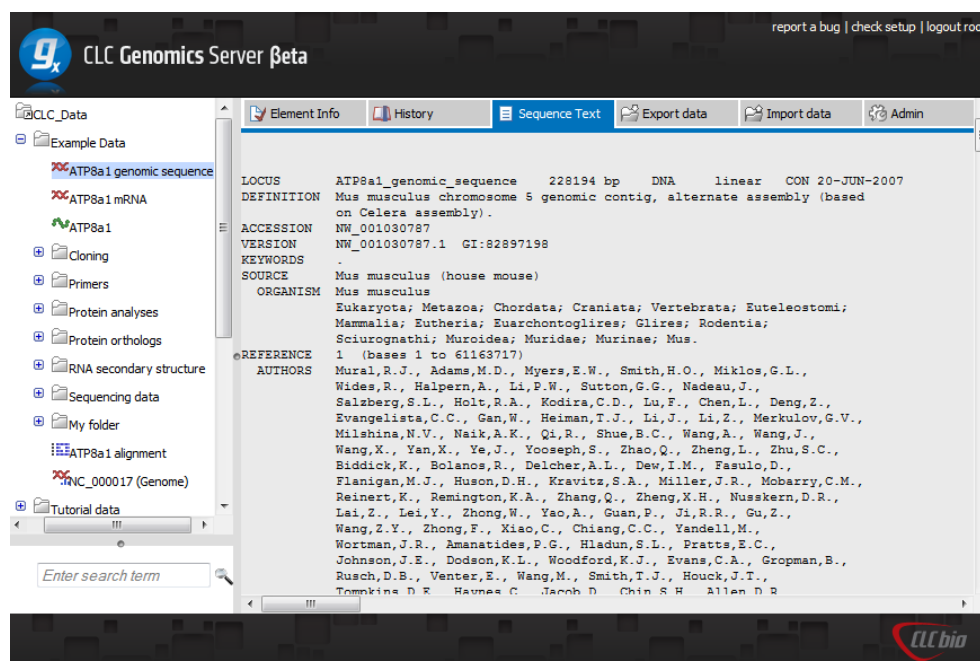An example of a protein sequence in the text view is shown in figure 3.2.

Figure 3.2: *Inspecting the text view of a protein sequence.*

Note that these views are a subset of the views that you find in the CLC Workbench.

3.2   Import and export

It is possible to import from and export to the server If you wish to import data from the server, click **Import** (🖼) and select the relevant data. Leave the file import format to 'Automatic' and press the 'Import File' button (Figure 3.3).

You can also put data into the import/export directories: Select the data you wish to export and click **Export** (🖼). Next, tick 'Save on server' and select the folder where you want the data to be (see figure 3.4).

Click the button labeled **Export**. The server will automatically recognize the file format and interpret the file.
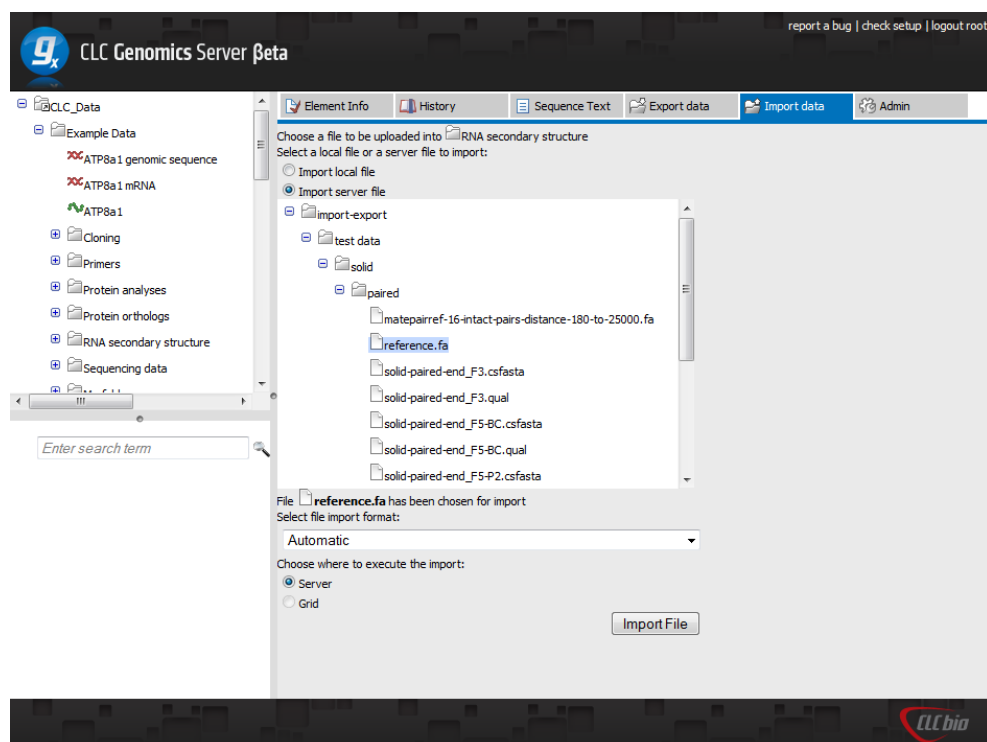
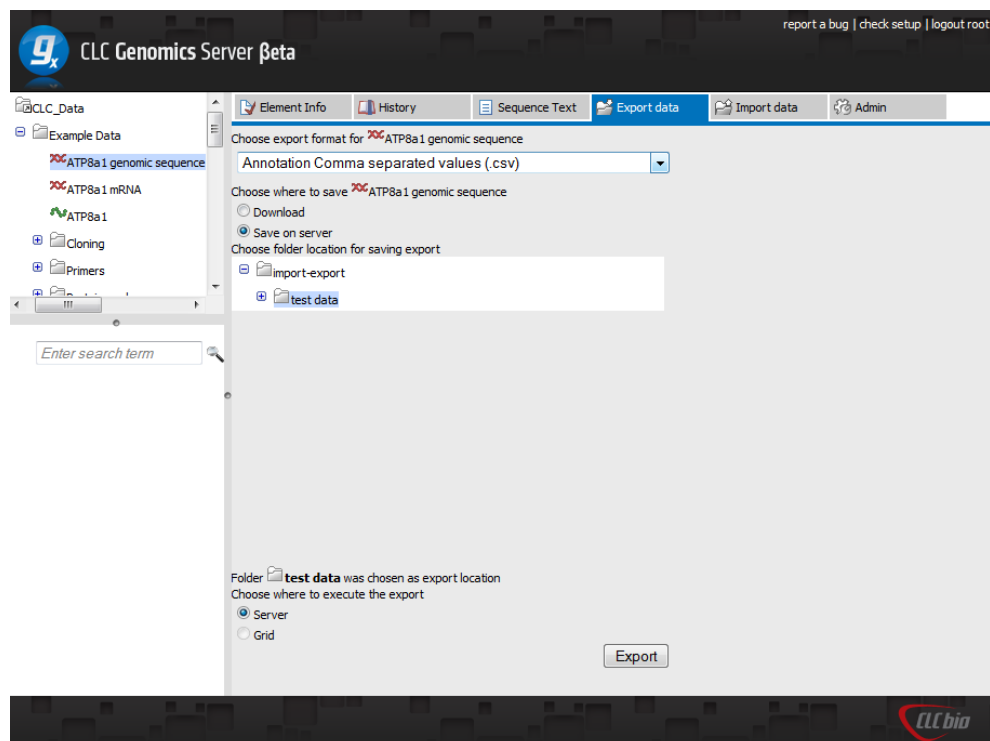Figure 3.3: *Importing a sequence from the server.*

Figure 3.4: *Exporting sequences to the server.*